Praktikum z ekonometrie

VŠE Praha

Tomáš Formánek

Block 4 – Linear mixed effect models – Outline

Introduction

LME generalization of a linear regression model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\omega}, \qquad \boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{G}), \qquad \boldsymbol{\omega} \sim N(\boldsymbol{0}, \boldsymbol{R}),$$

where

$\boldsymbol{y}$ is a vector of dependent variable observations,

$\boldsymbol{X}$ is a $(N \times p)$ matrix of N observations and p predictor variables,

$\boldsymbol{\beta}$ is a vector of "fixed-effects" regression coefficients,

$\boldsymbol{Z}$ is a $(N \times s)$ design matrix for the s "random effects" $\boldsymbol{u}$ that are complementary to $\boldsymbol{\beta}$,

$\boldsymbol{\omega}$ is the error term,

$\boldsymbol{u}$ and $\boldsymbol{\omega}$ are assumed normally distributed and mutually independent, with variance-covariance matrices $\boldsymbol{G}$ and $\boldsymbol{R}$ respectively.

Introduction

LME generalization of a linear regression model – example:

$$y_i = \beta_0 + \beta_1 x_i + \omega_i.$$

If observations $i = 1, \ldots, N$ are organized into $j = 1, \ldots, J$ relevant groups, and assuming there are random effects on both β-coefficients, we may generalize the model into a LME (after re-arranging):

$$y_{ij} = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j})x_{ij} + \omega_{ij},$$

where the combined ij subscript refers to an ith observation that belongs to a jth group.

In the LME model, u_{0j} and u_{1j} are stochastic deviations from β-coefficients that are associated with a particular jth group. While random effects may look like model coefficients, we are only interested in estimating their variances.

Introduction

Linear mixed effect model (LME) – data types:

- **Longitudinal data:** repeated measurements are performed on each individual unit. Several units are sampled. Number of observations may differ across units.

  y_{ti} - observation at time t for i-th individual.
  y_{ij} - ith observation of jth individual (if time aspect not relevant).

- **Hierarchical data structures:** data with two or more groups/levels of observations. Number of observations may differ across units.

  y_{ij} - observation for i-th company within j-th region.
  y_{ij} - observation for i-th student within j-th class.

- **Combined:** We can group observations at three levels (or more):

  y_{tij} - measurement at time period t, admin. region i within state j.

- Note how indices are ordered (left to right) from individual to highest level of aggregation. (alternative orderings exist in literature).

Longitudinal data

- N individual CS units are followed over time.

- The observation set $\{y_{ti}, \boldsymbol{x}_{ti}\}$ denotes some ith individual observed at a time period t. The number of observations in time may differ among CS units and observations may occur at different time points.

  Example: For a medical study, we measure child's weight (plus other data) at birth and repeatedly over a period of one year. For some y_{ti} observation, index t denotes days from birth. Due to doctor visit scheduling, children are weighted at different t "values". Typically, the number of doctor visits (observations) differs across children. Children in the study are born on different dates (say, Jan 2015 - Oct 2019).

  Example extends easily to economic environment
  (we can follow newly founded companies, etc.).

Hierarchical data structures

- Nested/hierarchical structure of the LME model:
    - Individual units i (Level 1) are nested
    - within j groups (Level 2) with group-specific observation sizes n_j.
- One or more coefficient(s) can vary across groups ("random effects).

LME: Longitudinal vs. hierarchical data structures:

- Essentially, the same nesting/hierarchical framework applies to longitudinal data and their LME-based analysis:
    - Observations at time t (Level 1) are nested
    - within j individual units (Level 2).
    - If appropriate, individual units can be nested in groups (Level 3) ...

Introduction

- Mixed models are called "mixed", because the β-coefficients are a mix of fixed parameters and random variables
- Terms "fixed" and "random" have specific meaning for LMEs:
  - A fixed coefficient is an unknown constant to be estimated.
  - A random coefficient varies from "group" to "group".
    By "group", we mean Level 2 aggregation, if data have 2 levels.
    - coefficients vary among schools (Level 2), not within school.
    - coeffs. vary across individuals (Level 2), not over time (Level 1).
- LME models can have some added complexity:
  - Multiple levels of nesting
  - Crossed random effects
  - Correlations between different random coefficients.
- Random coefficients are not estimated, but they can be predicted.

LME model example

- Data:
  London Education Authority Junior School Project dataset,
  - we have 887 students (i) in 48 different schools (j),
  - we want to predict 5th-year math scores.

- We may start by ignoring the school grouping and any possible regressors – we have a trivial model (*single-mean* model):

  $$\texttt{math5}_{ij} = \beta_0 + \varepsilon_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, M, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

  where $M = 48$ and n_j differ among schools, $\texttt{math5}_{ij}$ is the observed math score of i-th student at school j, β_0 is the mean math score across our population (being sampled) and ε_{ij} is the individual deviation from overall mean.

  Population mean math score & the variance of ε are estimated by taking their sample counterparts. Any "school effect" is ignored.

LME model

- The school effect (differences among schools) may be incorporated in the model by allowing the mean of each school to be represented by a separate parameter (*fixed effect*)

$$\texttt{math5}_{ij} = \beta_{0j} + \varepsilon_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, M, \quad \varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$$

where β_{0j} is the school-specific mean math score and ε_{ij} is the individual deviation from the school-specific mean.

  - `R` syntax: `lm(math5 ~ School-1, data=...)`
    $\Rightarrow$ $M = 48$ school-specific intercepts are estimated.
  - Using the terminology of LME, β_{0j} are fixed. Hence:
    - **Estimated intercepts only model (refer to) the specific sample** of schools, while -usually- the main interest is in the population from which the sample was drawn.
    - OLS regression does not provide an estimate of the between-school variability, which is also of central interest.

LME model

- *Random effects* approach: LME model can solve the above problem by treating school effects as random variations around population mean.

- Ordinary model (with *fixed effects*) can be reparametrized as:

  $y_{ij} = \beta_{0j} + \varepsilon_{ij}$

  $y_{ij} = \beta_0 + (\beta_{0j} - \beta_0) + \varepsilon_{ij},$

  Random effect: $u_{0j} = \beta_{0j} - \beta_0$ is the school-specific deviation from overall mean β_0. It can be used to replace the the *fixed effect* β_{0j} :

  $u_{0j} = \beta_{0j} - \beta_0 \quad \Rightarrow \quad \beta_{0j} = \beta_0 + u_{0j}$. Hence:

  $y_{ij} = \beta_0 + u_{0j} + \varepsilon_{ij}.$

- u_{0j} is a random variable, specific for the j-th school, with zero mean and unknown variance σ_u^2.

  u_{0j} is a *random effect*, associated with the particular sample units (schools are selected at random from the population).

LME model

- LME model with *random effects* (on the intercept) is given as:

$$y_{ij} = \beta_0 + u_{0j} + \varepsilon_{ij}, \qquad u_{0j} \sim N(0, \sigma_u^2), \qquad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2),$$

and we assume u_{0j} are *iid* and independent from ε_{ij}.

  - Observations within the same school share the same random effect u_{0j}, hence are positively "correlated" with ICC $= \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2)$ (see ICC discussion on a separate slide).
  - This *random effects* model has three parameters: β_0, σ_u^2 and σ_ε^2. (regardless of M, the number of schools).
  - Note that the *random effect* u_{0j} "looks like" a coefficient, but we are only interested in estimating σ_u^2.
  - However, upon observed data (and estimated model), we do make predictions using fitted values of $\hat{u}_j$.

LME model

- Exogenous regressors can be used in LMEs (like in LRMs).
  For example, `math5` grades depend on `math3` (3rd year grades).

  $\texttt{math5}_{ij} = (\beta_0 + u_{0j}) + \beta_1\,\texttt{math3}_{ij} + \varepsilon_{ij}$

  alternative notation:

  $\texttt{math5}_{ij} = \beta_0 + \beta_1\,\texttt{math3}_{ij} + u_{0j} + \varepsilon_{ij}$

  alternative notation:

  Level 1: $\quad \texttt{math5}_{ij} = \beta_{0j} + \beta_1\,\texttt{math3}_{ij} + \varepsilon_{ij}$
  Level 2: $\quad \beta_{0j} = \beta_0 + u_{0j}$

  - Intercept has a random effect, given the u_{0j} element.
  - Slope of the regression line for each school is fixed at β_1.
    ... `math3` has a *fixed effect*.

LME model: ICC

- **ICC:** Intra class correlation in a LME regression model:

$$\text{ICC} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}$$

  - Describes how strongly units in the same group are "correlated".
  - While interpreted as a type of correlation, ICC operates on groups, rather than paired observations.
  - See link for relation between ICC and actual correlation.

LME model: ICC

- **ICC:** Intra class correlation in a LME regression model:

$$\text{ICC} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}$$

- Example: $\texttt{math5}_{ij} = \beta_0 + \beta_1\,\texttt{math3}_{ij} + u_{0j} + \varepsilon_{ij},$

  where $\sigma_u^2 = \text{var}(u_{0j})$ and $\sigma_\varepsilon^2 = \text{var}(\varepsilon_{ij})$.

- ICC: “correlation” between `math5` observations (randomly chosen) within a given school.

- ICC has another useful interpretation: Say, ICC $= 0.6$ in our $\texttt{math5}_{ij}$ example. Hence, differences between schools explain 60% of “remaining” variance – i.e. after the variance explained by fixed effects (i.e. by $\texttt{math3}_{ij}$) is subtracted.

LME model: random effects on intercept and slope

- If teaching is different from school to school, it would make sense to have different slopes for each of the schools.

  Instead of using *fixed effects* on slopes (interaction terms `math3:School`), we use random slopes: $u_{1j} = \beta_{1j} - \beta_1$.

  $$\texttt{math5}_{ij} = (\beta_0 + u_{0j}) + (\beta_1\, \texttt{math3}_{ij} + u_{1j}\, \texttt{math3}_{ij}) + \varepsilon_{ij},$$

  alternative notation:

  $$\texttt{math5}_{ij} = \underbrace{\beta_0 + \beta_1\, \texttt{math3}_{ij}}_{\textit{fixed}} + \underbrace{u_{0j} + u_{1j}\, \texttt{math3}_{ij}}_{\textit{random}} + \varepsilon_{ij},$$

  - We can test whether this extra complexity is justified.
  - u_{0j} and u_{1j} are often correlated, their independence can be tested.
  - Fitted values of $\texttt{math5}_{ij}$ can be produced, along with $\hat{u}_{0j}$ and $\hat{u}_{1j}$.

LME: matrix form re-visited

- LME model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon} \qquad \boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{G}) \qquad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{R}),$$

  where:
  $\boldsymbol{X}$ is a $(n \times k)$ matrix, k is the number of *fixed effects*,
  $\boldsymbol{Z}$ is a $(n \times p)$ matrix, p is the number of *random effects*,
  $\boldsymbol{G}$ is a $(p \times p)$ variance-covariance matrix of the *random effects*,
  $\boldsymbol{R}$ is a $(n \times n)$ variance-covariance matrix of errors.

  - Independence between $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$ is assumed,
  - Often, $\boldsymbol{R} = \sigma_{\varepsilon}^2 \boldsymbol{I}_n$ is assumed,
  - $\boldsymbol{G}$ is diagonal if *random effects* are mutually independent.

- Estimation: MLE, RMLE, penalized least squares
  `https://www.jstatsoft.org/article/view/v067i01/0`

LME: R syntax

For a single random effect (on the intercept):

- `{nlme}` package:

  `lme( y ~ x + z, random = ~ 1 | g, data = df )`

- `{lme4}` package:

  `lmer( y ~ x + z + ( 1 | g ), data = df )`

- where `y` is the response variable with predictors `x` and `z`, and grouping factor variable `g`.

LME: R syntax

For random effects on the intercept and `x`:

- `{nlme}` package:

```
lme( y ~ x + z, random = ~ 1 + x | g, data = df )
lme( y ~ x + z, random =     ~ x | g, data = df )
```

- `{lme4}` package:

```
lmer( y ~ x + z + ( 1 + x | g ), data = df )
lmer( y ~ x + z +     ( x | g ), data = df )
```

LME: R syntax

For uncorrelated random effects on the intercept and x:

- {nlme} package:

```
lme( y ~ x + z, random = list ( g = pdDiag ( ~ x ) ) ,
     data = df )
```

- {lme4} package:

```
lmer( y ~ x + z + ( 1 | g ) + (0 + x | g) , data = df )
lmer( y ~ x + z + ( x || g ) , data = df )
```

Complex LME models - brief outline

Different types of LME models exist:

- LME models with (multilevel) nested effects,
- LME models with crossed effects,
- Complex behavior of the error term in LME models can be addressed (heteroscedasticity and serial correlation).
- LME models with non-Gaussian dependent variables (binary, Poisson, etc.).

LME models with multilevel nested effects

Multi-level model example: For 17 years, we follow a total of 86 individual states organized within 9 "global-level" regions (e.g. South America, Europe, Middle East, etc.).

- $\texttt{GDP}_{tij}$ represents individual GDP per capita measurements for:
  t-th time period, e.g. with values ($t = 2000, \dots, 2016$).
  i-th state nested within region j ($i = 1, \dots, M_j$),
  j-th region ($j = 1, \dots, 9$),
- We fit `GDP` as a function of productivity `P` and unemployment `U`.
  States are nested in regions, we have 2 levels of random intercepts:
  $u_{0i(j)}$ for each state (within a region),
  v_{0j} for the regions,
  random slopes can be added as well.
- $\texttt{GDP}_{tij} = \beta_0 + \beta_1 \, \texttt{P}_{tij} + \beta_2 \, \texttt{U}_{tij} + u_{0i(j)} + v_{0j} + \varepsilon_{tij}.$

LME: R syntax

For intercept varying among `g1` and `g2` within `g1`:
For intercept & `x` varying among `g1` and `g2` within `g1`:

- `{nlme}` package:

```
lme( y ~ x + z, random = ~ 1 | g 1 / g 2 , data = df )
lme( y ~ x + z, random = ~ x | g 1 / g 2 , data = df )
```

- `{lme4}` package:

```
lmer( y ~ x + z +( 1 | g 1 / g 2 ), data = df )
lmer( y ~ x + z +( x | g 1 / g 2 ), data = df )
```

LME models with crossed random effects

Crossed *random effects* example:

- Grunfeld (1958) analyzed data on 10 large U.S. corporations, collected annually from 1935 to 1954 to investigate how investment `I` depends on market value `M` and capital stock `C`.
- Here, we want *random effects* for a given firm and year.

  We want the year effect to be the same across all firms, i.e. not nested within firms.
- $\mathtt{I}_{ti} = \beta_0 + \beta_1 \, \mathtt{M}_{ti} + \beta_2 \, \mathtt{C}_{ti} + u_{0i} + v_{0t} + \varepsilon_{ti}.$

  where $i = 1, \ldots, 10$ and
  firms are followed over $t = 1, \ldots, 20$ years.
  (the usual "*it*" index ordering can be used as well)

LME: R syntax

For intercept varying among `g1` and `g2`

- `{lme4}` package:

  ```
  lmer( y ~ x + z + ( 1 | g 1 ) + ( 1 | g 2 ), data = df )
  ```

LME with heteroscedasticity and serial correlation

`{nlme}` package:

- Heteroscedastic residual variance at level 1:

```
lme( y ~ x + z, random = ~ 1 | g ,
          weights = varIdent ( form = ~ 1 | g ) , data = df )
```

- Autoregressive `ar(1)` residuals:

```
lme( y ~ x + z, random = ~ time | g ,
          correlation = corAR1 ( ) , data = df )
```

- General residuals (HAC estimation):

```
lme( y ~ x + z, random = ~ time | g ,
          weights = varIdent ( form = ~ 1 | time ) ,
          correlation = corAR1 ( ) , data = df )
```

LME models – references, R packages

- {lme4} package
  https://www.jstatsoft.org/article/view/v067i01/0
- {nlme} package
  https://cran.r-project.org/web/packages/nlme/nlme.pdf
- https://www.r-bloggers.com/2017/12/linear-mixed-effect-models-in-r/
- https://rpsychologist.com/r-guide-longitudinal-lme-lmer
- Finch, Bolin, Kelley: Multilevel Modeling Using R (2014).